Fuzzy sets and social research

Ragin, C.C.; Pennings, P.J.M.

published in
Sociological Methods and Research
2005

DOI (link to publisher)
10.1177/0049124105274499

document version
Publisher's PDF, also known as Version of record

citation for published version (APA)
Ragin, C. C., & Pennings, P. J. M. (2005). Fuzzy sets and social research. *Sociological Methods and Research*, *33*(4), 423-430. https://doi.org/10.1177/0049124105274499

Sociological Methods & Research

http://smr.sagepub.com/

Fuzzy Sets and Social Research

Charles C. Ragin and Paul Pennings

Sociological Methods & Research 2005 33: 423

DOI: 10.1177/0049124105274499

The online version of this article can be found at:

http://smr.sagepub.com/content/33/4/423

Published by:

SAGE

http://www.sagepublications.com

Fuzzy Sets and Social Research

CHARLES C. RAGIN
University of Arizona, Tucson

PAUL PENNINGS
Vrije Universiteit Amsterdam

This special issue of *Sociological Methods & Research* presents four in-depth methodological discussions of the use of fuzzy sets in social research. They have in common that they confront and compare fuzzy set methods with mainstream techniques. These contributions should not be read as introductions to fuzzy set analysis (see Smithson 1987; Ragin 2000) but as attempts to validate this new methodology and demonstrate some of its strengths by comparing it with established approaches.

In brief, fuzzy sets extend Boolean or "crisp" sets by permitting membership scores in the interval between 0 and 1. With crisp sets, cases are perceived only as members or nonmembers of a set. The problem is that many core concepts in social research are best understood as graded sets. Examples include such dichotomies as coordinated versus uncoordinated economies, national versus international politics, the public versus the private sector, states versus markets, consensus versus majoritarian systems, democratic versus nondemocratic, federal versus nonfederal, employed versus unemployed, male versus female, high versus low, established versus nonestablished, rich versus poor, and so on (see Pennings 2003). At a theoretical level, most researchers are fully aware of the problematic aspects of using these concepts as simple dichotomies. But this awareness has not been translated into the application of methodologies that are fully equipped to study diversity and complexity in a set-theoretic manner.

Fuzzy sets can help social scientists conceptualize social and political phenomena as sets with imprecise boundaries between

SOCIOLOGICAL METHODS & RESEARCH, Vol. 33, No. 4, May 2005 423-430
DOI: 10.1177/0049124105274499

membership and nonmembership (Klir and Yuan 1995: 4). For example, a country (e.g., the United States) might receive a membership score of 1 (full membership) in the set of rich countries but a score of only .8 (less than full membership) in the set of democratic countries. The basic idea behind fuzzy sets is to permit the scaling of membership scores and thus allow partial or fuzzy membership. A membership score of 1 indicates full membership in a set, scores close to 1 (e.g., .8 or .9) indicate strong but not quite full membership in a set, scores less than .5 but greater than 0 (e.g., .2 and .3) indicate that objects are more "out" than "in" but still weak members of a set, and a score of 0 indicates full nonmembership in a set. Thus, fuzzy sets combine qualitative and quantitative assessment: 1 and 0 are qualitative assignments ("fully in" and "fully out," respectively); values between 0 and 1 indicate partial membership. The .5 score is also qualitatively anchored, for it indicates the point of maximum ambiguity (fuzziness) in the assessment of whether a case is more "in" or "out" of a set.

Fuzzy membership scores address the varying degree to which different cases belong to sets, not how cases rank relative to each other on dimensions of open-ended variation. Thus, fuzzy sets pinpoint qualitative states while at the same time assessing varying degrees of membership between full inclusion and full exclusion. In this sense, a fuzzy set can be seen as a continuous variable that has been purposefully calibrated to indicate degree of membership in a defined set. Such calibration is possible only through the use of theoretical and substantive knowledge, which is essential to the specification of the three qualitative breakpoints: full membership (1), full nonmembership (0), and the point of maximum ambiguity regarding membership (.5). These qualitative anchors are used to distinguish between relevant and irrelevant variation. For example, the substantial variation in gross national product (GNP) per capita among the unambiguously rich countries is not relevant to membership in the set of rich countries, at least from the perspective of fuzzy sets. If a country is unambiguously rich, then it is accorded full membership, a score of 1. Similarly, variation in GNP per capita among the unambiguously not-rich countries is also irrelevant to membership in the set of rich countries (membership = 0). Thus, in research using fuzzy sets, it is not enough to develop scales that show the relative positions of cases (e.g., as with a conventional interval-scale measure of wealth such as

GNP per capita). It is also necessary to use qualitative anchors to map the links between specific scores on continuous variables (e.g., GNP per capita) and degree of membership in a well-specified set (e.g., degree of membership in the set of rich countries).

Fuzzy sets retain almost all the essential mathematical properties of crisp sets and thus enable researchers to model complex and diverse constellations of case aspects and to assess set-theoretic relations. Set-theoretic relations are central to social theory and to empirical research, but they are largely beyond the reach of conventional quantitative techniques (the contributions of Goertz and Mahoney on two-level theories and of Katz, vom Hau, and Mahoney in this special issue are demonstrations of this). For example, fuzzy sets permit the identification of necessary and sufficient conditions by means of the subset principle: A condition can be interpreted as necessary if its membership scores are consistently greater than degree of membership in the outcome (the condition is a superset of the outcome). A condition can be interpreted as sufficient if its memberships scores are consistently less than degree of membership in the outcome (the condition is a subset of the outcome). It is surprising to us that fuzzy sets are not used more often in social science, given that almost all verbal theory is formulated in terms of set-theoretic relations and almost all social science concepts can be usefully formulated as fuzzy sets.

The major claims being advanced in this special issue about the applicability of fuzzy sets to social research include the following:

- Fuzzy sets permit a more nuanced representation of categorical concepts by permitting degrees of membership in sets rather than binary in-or-out membership.
- Fuzzy sets can be used to address both diversity and ambiguity in a systematic manner, through set calibration and set-theoretic relations.
- Most verbal theory in the social sciences is formulated explicitly in set-theoretic terms. The fuzzy set approach provides a faithful translation of such theory.
- Fuzzy sets enable researchers to evaluate set-theoretic relationships such as intersection and inclusion and, thereby, necessity and sufficiency. Set-theoretic relationships are very difficult to evaluate using conventional approaches such as the general linear model.

Although we argue that fuzzy sets can substantially augment researchers' methodological tool boxes, we do not want to make exaggerated claims. Researchers do need to be aware of crucial issues and choices involved in applying fuzzy sets. Accordingly, the object of the contributions of Michael Smithson (2005 [this issue]) and Jay Verkuilen (2005 [this issue]) is to introduce researchers to some of these issues and point the way toward possible solutions.

Smithson's (2005) contribution links fuzzy set methods to mainstream techniques. He takes the position that fuzzy sets and statistics work better together than separately. His contribution begins by establishing an important link between fuzzy sets and cumulative distribution functions (cdfs), from the observation that any membership function may be represented as a series of nested sets. It turns out that for any equal-interval fuzzy membership function, the sum of the cdf for any sample is the sum of the degrees of membership in that sample. Moreover, other simple functions of the cdf are directly related to the variance in membership and covariance for two fuzzy sets. These linkages are important because they relate fuzzy set methods directly to conventional statistics, as well as providing alternative interpretations of them.

He then applies the cdf approach to clarifying fuzzy set intersection, union, inclusion, necessity, and sufficiency. More specifically, he shows how to use well-known statistical tests (e.g., chi-square and Kolmogorov-Smirnov) to compare the extent to which two fuzzy sets intersect or include one another against specific alternatives (e.g., statistical independence or particular inclusion rates). In addition, Smithson shows how to use simple statistical tests (e.g., chi-square) and confidence intervals to evaluate fuzzy necessity and sufficiency without making strong assumptions about membership scales. Where stronger measurement assumptions are permitted, Smithson demonstrates how more focused questions regarding necessity and sufficiency may be addressed. He then compares the characterizations obtained from correlation and regression with those from fuzzy necessity/sufficiency and develops criteria for determining when one would reasonably be preferred over the other. Finally, he briefly surveys statistical techniques that are potentially compatible with fuzzy set analysis, such as profile analysis, (fuzzy) clustering, and certain models for ordinal categorical data.

Verkuilen (2005) argues in his contribution that uncertainty in social science is unavoidable due to two basic sources: ambiguity from multiple meanings for background concepts and vagueness from lack of firm boundaries. Fuzzy sets can help manage vagueness but must presume that concepts are not ambiguous. Unfortunately, membership assignment is one of the weak points of much of the fuzzy set literature. The requirements are validity (measures must actually measure the concept at hand and not be biased) and reliability (measures must be reasonably reproducible if repeated). In addition, we must meet some required mathematical structure (unidimensional continuous latent variable in the unit interval, with a minimal level of ordinality and natural endpoints). Verkuilen argues that there is no "magic bullet" for these problems. Only intellectual discipline and an effort to make the rules used in assignment as clear as possible will help. It is important to be flexible and not fall into intellectual "defaults." The membership assignment task is completed in four steps. First, concepts need to be systematized by means of specification and defining the negation. Second, indicators should be chosen. Third, membership is to be assigned by finding a transformation that takes the indicators into the unit interval. Finally, measures should be validated.

The methods of assignment can be divided into direct versus indirect assignment by one or more than one indicator. Verkuilen (2005) provides a practical guide with two detailed examples, which illustrate assignment using survey data and assignment by scaling multiple indicators using cross-national data. Both examples were chosen because the authors posit a necessary condition hypothesis that would be testable in a fuzzy set framework.

The contribution of Goertz and Mahoney (2005 [this issue]) focuses on "two-level theories" and fuzzy set logic. Their goal is twofold: (1) to introduce the theoretical structure of two-level models and (2) to examine the methodological and empirical issues involved in the application of fuzzy set methods to two-level theories. To introduce and exemplify two-level theories, they use Skocpol's (1979) *States and Social Revolutions* as their main example. Despite the extensive examination and discussion that this work has received, their analysis provides for the first time a succinct and accurate portrayal of the structure of its theory.

Two-level models embody two types of theories. The theory at Level 1 focuses on the causes of the main dependent variable of interest (in this case, social revolution). By contrast, the theory at

Level 2 seeks to explain the Level 1 causes themselves; that is, the Level 1 causal conditions are the effects of the Level 2 causal conditions. The authors argue that the theoretical and methodological issues vary quite significantly from one level to the other. Level 1 is the core of two-level models because it contains the explanation for the main dependent variable of interest. In the case of Skocpol (1979), this level entails the claim that peasant revolt and state breakdown are individually necessary and jointly sufficient for social revolution. The authors focus on theories that have a similar causal structure at Level 1, that is, theories in which the causes are necessary and jointly sufficient. This is quite different from the logic at Level 2, which is an argument about several variables that are individually sufficient.

The discussion proceeds in two parts. Part 1 formally describes the structure of the two-level theories considered and shows how this theoretical structure applies to Skocpol's (1979) work and a variety of other prominent works. Part 2 turns the attention to the use of fuzzy set logic as a tool for assessing two-level theories by analyzing Skocpol's work on revolutions. They conclude that fuzzy set methods are helpful for analyzing two-level theories because they are explicitly designed to think about causation in terms of necessary and sufficient conditions. In addition, fuzzy set methods offer logical devices for conceptualizing the relationship between Level 1 and Level 2 variables. At the same time, however, their evaluation of two-level theories suggests a number of refinements to fuzzy set methodology. For one thing, they illustrate the difficulty that these methods might have evaluating two-level theories if one is not clear about the structure of these theories from the outset. Hence, analysts should take care to consider the overall structure of a theory before evaluating its variables using fuzzy set techniques.

Finally, the contribution of Katz et al. (2005 [this issue]) evaluates the relative strengths and weaknesses of fuzzy set analysis and regression analysis for explaining the "great reversal" in Spanish America. From 1750 to 1900, the most marginal colonial territories often became the region's wealthiest countries, whereas the most central colonial territories often became the region's poorest countries. To explain this reversal, five competing hypotheses are tested using both regression and fuzzy set methods. The fuzzy set analysis reaches substantively important conclusions, finding that strong liberal factions are probabilistically necessary for economic development, and dense indigenous populations are probabilistically necessary for social underdevelopment. By contrast, the regression analysis generates findings that are not meaningful. These contrasting results are traced to differences in how the methodologies understand causation. The article concludes that fuzzy set analysis and regression analysis operate in different "causal universes" and that greater attention should be granted to the causal universe occupied by fuzzy set analysis.

The four contributions illustrate that fuzzy sets are useful for both qualitatively and quantitatively orientated researchers and that the qualities of fuzzy set theory warrant a more frequent use and application in the social sciences. Although the fuzzy set approach originated in the mid-1960s, it took a long time before it was introduced into the social sciences, and even today, few applications have been made (although the number is steadily rising). The contributions in this special issue show that fuzzy set logic has the capacity to assign partial membership to cases of sets, which has important implications for the methodological treatment of many central concepts in the social sciences. In addition, the comparison with mainstream techniques shows that fuzzy sets offer many unique possibilities for finding causal patterns, which are not addressed by the established approaches. Finally, the contributions make clear that there is not one way to apply fuzzy sets. In fact, the methodology of fuzzy set is very diverse in itself, which is reflected, for example, by the many ways in which fuzzy set assignments can be derived. This special issue therefore illustrates that the added value of fuzzy sets is no longer a claim made by a particular small group of comparativists, but it has materialized into firm and unique findings, which warrant further investigation and application.

REFERENCES

Goertz, Gary, and James Mahoney. 2005. "Two-Level Theories and Fuzzy-Set Analysis." *Sociological Methods & Research* 33 (4): 497-538.

Katz, Aaron, Matthias vom Hau, and James Mahoney. 2005. "Explaining the Great Reversal in Spanish America: Fuzzy-Set Analysis Versus Regression Analysis." *Sociological Methods & Research* 33 (4): 539-573.

Klir, G. A., and B. Yuan. 1995. *Fuzzy Sets and Fuzzy Logic*. New York: Academic Press.

Pennings, P. 2003. "Beyond Dichotomous Explanations: Explaining Constitutional Control of the Executive With Fuzzy-Sets." *European Journal of Political Science* 42 (4): 541-67.

Ragin, C. C. 2000. *Fuzzy-Set Social Science*. Chicago: University of Chicago Press.

Skocpol, Theda. 1979. *States and Social Revolutions*. Cambridge, UK: Cambridge University Press.

Smithson, Michael. 1987. *Fuzzy Set Analysis for Behavioral and Social Sciences*. New York: Springer-Verlag.

———. 2005. "Fuzzy Set Inclusion: Linking Fuzzy Set Methods With Mainstream Techniques." *Sociological Methods & Research* 33 (4): 431-461.

Verkuilen, Jay. 2005. "Assigning Membership in a Fuzzy Set Analysis." *Sociological Methods & Research* 33 (4): 462-496.

Charles C. Ragin holds a joint appointment as professor of sociology and political science at the University of Arizona, Tucson. His most recent book, Fuzzy-Set Social Science, *was published in 2000. He has developed two software packages for set-theoretic analysis of social data, Qualitative Comparative Analysis (QCA) and Fuzzy-Set/Qualitative Comparative Analysis (fsQCA). In 2001, he was awarded the Donald Campbell Award for Methodological Innovation by the Policy Studies Organization.*

Paul Pennings is an associate professor in the Department of Political Science at the Vrije Universiteit Amsterdam. He has published widely on comparative politics and methodology. His recent work includes Doing Research in Political Science: An Introduction to Comparative Methods and Statistics *(with Hans Keman and Jan Kleinnijenhuis; second edition, 2005).*